

Critiquing Recommenders for Public Taste Products

Pearl Pu
EPFL
BC 107 EPFL
CH-1015 Lausanne
Switzerland
pearl.pu@epfl.ch

Maoan Zhou
EPFL
BC 107 EPFL
CH-1015 Lausanne
Switzerland
maoan.zhou@epfl.ch

Sylvain Castagnos
EPFL
BC 145 EPFL
CH-1015 Lausanne
Switzerland
sylvain.castagnos@epfl.ch

ABSTRACT

Critiquing-based recommenders do not require users to state all of their preferences upfront or rate a set of previously experienced products. Compared to other types of recommenders, they require relatively little user effort, especially initially, despite potential accuracy problems. On the other hand, they rely on a set of critiques to elicit users feedback in order to improve accuracy. Thus the better the critiques are, the more accurately and efficiently the system becomes in generating its recommendations. This method has been successfully applied to high-involvement products. However, it was never tested on public taste products such as music, films, perfumes, fashion goods or wine. Indeed our initial trial adapting traditional critiquing methods to this new domain led to unsatisfactory results. This has motivated us to develop a novel approach named "editorial picked critiques" (EPC) that accounts for users' needs for popularity information, editorial suggestions, as well as their needs for personalization and diversity. Through an empirical study, we demonstrate that EPC presents a viable recommender approach and is superior on several dimensions to critiques generated by data mining methods.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and presentation]: User Interfaces - *evaluation/methodology, graphical user interfaces (GUI), user-centered design.*

General Terms

Design, Experimentation, Human Factors

Keywords

Critiquing based recommender system, compound critiques, public taste products, user acceptance of recommenders

1. INTRODUCTION

A critiquing-based recommender system simulates an artificial salesperson. At first, it recommends products based on a client's current preferences and then elicits her feedback in the form of critiques such as "I like something cheaper" or "I like something

with faster processor speed". The critiques were generated with the aim of collecting users' preferential feedback in order to generate more accurate and personalized recommendations in the next interaction cycle. The simplest form of critiquing is called unit critiquing [2]. For example, [CPU Speed: faster] is a unit critique over the CPU speed attribute of the proposed PC. To make the critiquing process more efficient, compound critiques can be dynamically generated to enable users to simultaneously make critiques on several attributes in a given interaction cycle [5].

McCarthy et al. [5] have proposed a method to generate compound critiques using the Apriori algorithm [1], which was originally used in a supermarket management system. Zhang and Pu [7] presented an approach to generate compound critiques for conversational recommenders with users' preference models based on the multi-attribute utility theory (MAUT). Independently, Chen and Pu [3] proposed a preference-based critique generation method to organize recommended products into several clusters of critiques.

Both [3] and [5] generate the critiques based on the structure of a product catalog using the Apriori algorithm. They function well for high-involvement products such as PCs, digital cameras, and travel planning where consumers are more motivated to spend a significant amount of time in choosing the right product and are more likely to rely on their own judgment in the selection process (hence the name high involvement). This is largely due to the fact that a bad decision can result in significant financial losses. However, such methods have not been applied to public taste goods, such as music, films, perfumes and wines, where users spend less time in choosing the products and are more likely to rely on public opinion or experts' advice to make decisions. For example, Wine.com, a well-recognized wine selling website, provides editorial picks by combining eight professional ratings on wines such as "90+ rated wines under \$20". ESPN, one of the most reputed sports broadcasters, has "Expert Picks" for matches like NFL to give professional opinions to spectators. The MSN shopping website¹ provides editorial information to help consumers make purchase decision in a large number of fields, such as furniture, fragrance, and clothing. Smith et al. [6] showed that users significantly prefer editorial recommendations, especially for utilitarian products.

The present research work investigates whether previous work on critiquing based recommendation methods can be directly applied to public taste products. The contribution lies in the development

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

RecSys '09, October 23–25, 2009, New York, New York, USA.

Copyright 2009 ACM 978-1-60558-435-5/09/10...\$10.00.

¹ Shopping.msn.com

and validation of a new critique generation algorithm, called editorial picked critiques, or EPC, which is more suitable for recommendations of public taste products. The new algorithm combines editorial opinions, popularity information and critiques which were generated by data mining techniques. In a comparative user study, we learned that EPC is better in all of the 12 dimensions that were used in the evaluation and 9 of them have p-values less than 0.05. Users' average rating for EPC's ability to show more attractive items is 21% higher, users' confidence for purchase is 29% higher and EPC is 2.42 times more preferred than a general recommender. These results help establish that EPC is not only a promising and viable recommender, but also one that significantly improves users' decision to purchase due to increased confidence.

2. PERFUME RECOMMENDER

2.1 System Overview

The application for editorial picked critiques (EPC) was implemented on our group's product brokering website for finding perfume, cosmetics, fashion goods, wines, etc. When a particular product catalog is selected, the website displays a page as shown in Figure 1. A multi-criteria filtering tool is located at the top of this page. Users can select any of the five attributes to view a subset of the perfumes. Alternatively, they can view the entire catalog by selecting a sorting criterion such as popularity (default), brand, and price.

Home » LATEST TREND » PERFUME (WOMEN)

2363 total results for PERFUME (WOMEN) European perfumes sold in retail stores worldwide

brand	price	quantity	category
Chanel (42)	Less than 30 USD (565)	Less than 40 ml (262)	Eau de Parfum (1071)
Estee Lauder (33)	30-50 USD (643)	40-80 ml (843)	Eau de Toilette (1188)
Calvin Klein (15)	50-80 USD (527)	80-120 ml (752)	
Gucci (6)	80-110 USD (185)	120-160 ml (48)	
Givenchy (11)	More than 110 USD (243)	More than 160 ml (22)	
More...			

Total results: 2363

currency: USD sort list by: Popularity show: 16 1 2 3 ...

Givenchy Very Irresistible Eau de Parfum, 2.5 oz. Givenchy. Around 77 USD. 73ml. A red carpet of roses illuminated by star anise and the ...

Kai Perfume Kai Eau de Parfum Spray, 3.4 oz. KooKai. Around 65 USD. Shipping March 5th.... What you've asked for ...

Romance by Ralph Lauren for Women 3.4 oz Eau de Parfum Spray, 100ml. RALPH LAUREN. Around 61.99 USD. The intimacy of Romance by Ralph Lauren. A sensual ...

Kai Perfume Kai Eau de Parfum Spray, 3.4 oz. KooKai. Around 45 USD. An intoxicating blend of tropical gardenia and white exotic ...

Angel by Thierry Mugler for Women 0.5 oz Eau de Parfum Spray Refillable (Decoded Box), 23ml. THIERRY MUGLER. Around 47.98 USD. Angel by Thierry Mugler ...

Miracle So Magic! by Lancome for Women 3.4 oz Eau de Parfum Spray, 100ml. LANCOME. Around 47.98 USD. A fresh, floral and vibrant fragrance. A daring ...

Figure 1. Product Listing Page with the Search Panel

When a user clicks on the image of a particular product, she will be directed to the product detail page (as shown in Figure 2) and presented with the product's detailed information, as well as the results of the editorial picked recommendations on the right hand side under "You may also like" in different categories.

2.2 Perfume Dataset

In our present research, we have chosen to work with data related to perfumes. They are relatively easy to obtain, they are public taste goods, people buy them more frequently than PCs and digital cameras, and they are sufficiently interesting and complex in terms of the number of features that describe them (brand, price, quantity, etc.) We have thus obtained more than 4000 perfume products with popularity information from a well known e-commerce website. A series of attributes for perfumes from the

Miracle So Magic! by Lancome for Women 3.4 oz Eau de Parfum Spray

LANCOME

Miracle So Magic! by Lancome for Women 3.4 oz Eau de Parfum Spray 100ml

Price: Around \$47.98

Features:

Quantity: 100 ml

Category: Eau de Parfum

Gender: Women

Bestselling rate: 9.5

Average user rate: 0

Source: Amazon

Add to shopping list

A fresh, floral and vibrant fragrance. A daring association of fresh clover leaf with flowery heart notes of wild rose and narcissus, to create a truly vibrant fragrance.

User Rating: ☆☆☆☆ / 0

Uninteresting ○ ○ ○ ○ I love it (Rate)

You may also like...

More popular and cheaper (1/1)

Kai Perfume Kai Perfume Oil \$ 45

An intoxicating blend of tropical gardenia and white exotic flowers. Previously packed in a roll-on vial...

Same brand and cheaper (1/6)

Attraction by Lancome for Women 1.7 oz Eau de Parfum Spray \$ 34.59

It begins with light, the inexplicable flash between a man and a woman. A look. A touch. A passionate emb...

Just as popular, but cheaper (1/6)

Green Tea Scent by Elizabeth Arden for Women 1.7 oz Eau de Parfum Natural Spray \$ 13.99

Evoke feelings of pure serenity. Clean and refreshing, with peaceful notes of Green Tea accord...

Same price range, just as popular (1/6)

Kai Perfume Kai Perfume Oil \$ 45

An intoxicating blend of tropical gardenia and white exotic flowers. Previously packed in a roll-on vial...

People who like this may also like (1/6)

Versace Signature by Gianni Versace for Women 3.4 oz Eau de Parfum Spray \$ 47.99

NULL...

Figure 2. Perfume detail page and editorial picked recommendations

resulting data was then put together, taking into account the brand, gender, price, quantity, category, popularity and description, and subdivided into 4 categories i.e. "Eau de Parfum", "Eau de Toilette", "Cologne" or "Aftershave"(only for men).

3. EDITORIAL PICKED CRITIQUES

Editorial picked critiques are generated by combining editorial opinions (e.g., from popular magazines), popularity information and critiques generated by the Apriori algorithm. In the beginning, we predefine some important unit critique categories that match users' attention and needs. These categories are put in a ranked list with respect to their influence and relevance to users' preferences. We then produce recommendations for each critique category based on a combined method of dynamic and preference based critiquing methods ([5] and [3] respectively).

3.1.1 Critique Categories

We identified five critique categories for public taste goods. Even though the current method was developed for perfume data, it can be generalized for fashion goods, wines, music, and films. Each category of critiques is engaged in one aspect of the products. Recommendations are produced from these categories in a top down order.

1. *Price-driven critiques.* Price is determined as the most important critique for our users.

2. *Popularity-driven critiques.* The second category is associated with popularity. It is interesting to ask why popularity is put above many other crucial factors. Kim and Chung [4] analyzed this question within their study and showed that "popularity can provide assurance to buyers, particularly when consumers evaluate products the features of which are not easily compared among alternatives (e.g. automobiles)". Assuming all other factors are equal, consumers reduce risk by purchasing popular models rather than unpopular ones. Second, popularity information, especially those gathered from e-commerce website with a significant amount of traffic, offers users a rare opportunity to

Table 1. Questions used to compare two recommender systems

ID	Questions
1	The items under "You may also like" are attractive.
2	The items under "You may also like" are educational.
3	The items under "You may also like" appeared to be good deal.
4	The items under "You may also like" appeared to be marketing material.
5	The items under "You may also like" influence my selection.
6	The items under "You may also like" will influence my future selection.
7	Names of the categories are useful and adequate.
8	I am satisfied with the overall quality of the interface.
9	I found the interface easy to use.
10	I would buy the perfume recommended to me, given the opportunity.
11	If it is a real website, I would use it in the future to find the perfume.
12	I would like to introduce this system to my friends.

observe what other people are buying. Third, quality associated with popularity will affect the perception of the brand, and this brand will consequently be more likely to be selected among other alternatives.

3. *Diversity-driven critiques.* We define the third category according to some common features that provide diversification for the product domain. For perfume data, it is quantity since the same perfume can exist in different sizes of bottles. Two quantity related critiques are proposed:

- (1) Same product, but less and cheaper.
- (2) Same brand, but less and cheaper.

For fashion items, we can use features such as color or occasion. For movies and files, genre can be used as the diversification variable.

4. *Similarity-driven critiques.* In this category, we try to propose something that is similar to the current product. The critique's name here is defined as "People who like this may also like ...",

which corresponds to content-based recommendation rather than social filtering. In the experiment that we present in the next section, we propose perfume that is similar to the current selected one according to their attribute distance. We look for items which have the smallest tradeoff differences in their attributes compared with the current product.

5. *Special recommendation.* Similar to an editor's special picks, items in this category aim at inspiring users to see more popular items, but those that are slightly more expensive. They provide educational value to users especially when they are not yet familiar with the domain.

Based on these unit critiques, we generated a set of five compound categories: "more popular and cheaper" or if this former category does not contain any real products we will use "more popular but more expensive", "same brand and cheaper" or "same brand but more expensive", "just as popular and cheaper", "same price range and just as popular", and "people who like this may also like".

Table 2. Average scores for each of the 12 evaluation questions

Questions	EPC	Apriori	P-Value
<i>Attractive recommender</i>	3.8636	3.1818	0.0041
<i>Educational recommender</i>	3.4545	3.1818	0.2482
<i>Good deal recommender</i>	3.8636	3.3636	0.0378
<i>Marketing material</i>	3.5909	3.4091	0.4626
<i>Influenced my selection</i>	3.6818	3.0455	0.0270
<i>Influence future selection</i>	3.7273	3.2273	0.0306
<i>Name of categories adequate</i>	3.6818	2.9091	0.0078
<i>Interface quality</i>	3.7727	3.2328	0.0179
<i>Interface easy to use</i>	3.6364	3.5000	0.1858
<i>Would buy given opportunities</i>	3.6818	2.8636	0.0010
<i>Would re-use</i>	3.5455	3.1818	0.0023
<i>Would introduce to friends</i>	3.4545	2.7273	0.0005

To validate whether these categories are attractive to users, we conducted a small user study with 7 volunteers with the help of an eye-tracking system. The results show that users indeed paid most attention to price and popularity attributes. They liked products that were "more popular and cheaper" the best, followed by the "same brand and cheaper", and "similar popularity and cheaper price". In addition, they showed interest in products from "People who like this may also like". Users also liked information about products with a higher price but superior popularity.

4. EXPERIMENT

To validate our work, we decided to conduct a user study to compare the newly developed algorithm, EPC, with a generalized version of both dynamic and preference based critiquing serving as the baseline. We will call the latter the Apriori method. A total of 22 participants took part in our experiment in a with-in subject setting. Thus every user experienced both EPC and Apriori in a randomized order. The user task was to find a perfume that she or he was ready to buy. Each participant was told that if she won the lottery of the experiment, she would obtain a 100 Swiss-Franc voucher to buy the selected perfume.

4.1 Evaluation Criteria

We designed a list of questions to measure users' subjective perception of the respective recommenders in terms of users' satisfaction of the recommendations (content assessment), the ease of use of the interfaces (interface assessment), and their intention to purchase recommended products and return to the website in the future (purchase and loyalty) (Table 1). Every question was responded on a 5-point Likert scale from "strongly disagree" to "strongly agree". At the end of the questionnaire, users were asked to give their preferences over the two systems: 1) which system was better at recommending perfume to you? and 2) which system did you prefer?

4.2 Experiment Results

Table 2 shows the average scores and p-values of subjective variables measured by the questionnaire for the two systems.

Figure 3 depicts the scores of each question used in the post study questionnaire. In all dimensions, scores for EPC are higher than those of Apriori and 9 of them have p-values less than 0.05. Scores for the question on purchase ("I would buy given

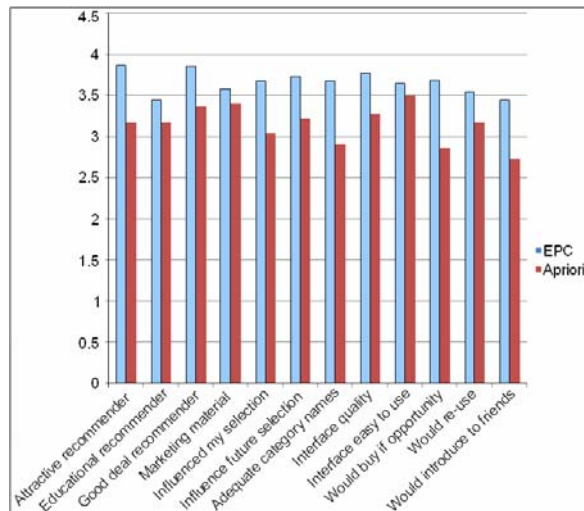


Figure 3. Average scores of each evaluation question

opportunities") is 29% higher for EPC than for Apriori. This result shows that users' purchasing decision is significantly improved by the EPC recommender. In addition, EPC is 21% more attractive, users prefer the names of the categories used in EPC and are more likely to reuse EPC again and introduce it to their friends, and EPC is 2.42 times more preferred over a general recommender.

5. CONCLUSION

The initial adaption of both dynamic and preference based critiquing recommenders to public taste data did not give satisfactory results in terms of user acceptance. The categories generated by both methods do not correspond to users' needs. In this paper, we presented a novel critiquing system, editorial picked critiquing, for recommending perfume which is a typical public taste product. This approach combines popularity scores and experts' choice of critiques. It overcomes some of the limitations of previously developed methods, which all rely on the Apriori data mining method to generate critiques. The advantages of EPC are highlighted and validated in a user experiment. More specifically, EPC can significantly improve the quality of recommendations, both in terms of recommendation content and the ease of use of the interface. In addition, the improved quality seems to inspire users' trust as well as their confidence in the system and help them make the purchase decision. In the future, we like to investigate how this method can be applied to other public taste goods such as wine, films, music, etc. in order to develop a set of design guidelines for the identification and testing of the critique categories.

6. ACKNOWLEDGMENTS

We thank the Swiss National Science Foundation for sponsoring the reported research work. We are grateful to the participants of our user studies for their patience and time. We also like to thank the reviewers for their suggestions to improve the clarity of the paper.

7. REFERENCES

- [1] Agrawal, R., Imielinski, T., and Swami, A. 1993. Mining association rules between sets of items in large databases. In Proc. of the 1993 ACM SIGMOD international conference on Management of data, 207-216
- [2] Burke, R.; Hammond, K.; and Young, B. 1997. The FindMe Approach to Assisted Browsing. IEEE Expert: Intelligent Systems and Their Applications 12(4):32-40.
- [3] Chen, L. and Pu, P. 2006. Evaluating Critiquing-based Recommender Agents. In Proceedings of Twenty-first National Conference on Artificial Intelligence (AAAI-06), pages 157-162, Boston, USA, July 16-20, 2006. □
- [4] Kim, C. and Chun, Y. 1997. Brand Popularity, Country Image and Market Share: an Empirical Study. Journal of International Business Studies 28(2), 361-386
- [5] Reilly, J., McCarthy, K., McGinty, L., and Smyth, B. 2004. Dynamic critiquing. In Proc. of the 7th European Conference on Case-Based Reasoning (ECCBR'04), 763-777.
- [6] Smith, D., Menon, S. and Sivakumar, K. 2005. Online Peer and Editorial Recommendations, Trust, and Choice in Virtual Markets. Journal of Interactive Marketing 19(3), 15-37.
- [7] Zhang, J. and Pu, P. 2006. A comparative study of compound critique generation in conversational recommender systems. In Proc. of the 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'06), 234-243.